

An elementary approach to Stochastic Differential Equations using the infinitesimals.

Vieri Benci*, Stefano Galatolo*, and Marco Ghimenti*

July 23, 2008

Contents

1	Introduction	1
2	The Alpha-Calculus	4
2.1	Basic notions of Alpha-Theory	4
2.2	Infinitesimally small and infinitely large numbers.	7
2.3	Some notions of infinitesimal calculus	8
3	Grid functions	10
3.1	Basic notions	10
3.2	The Ito formula	12
3.3	Distributions and grid functions	13
4	Stochastic differential equations	15
4.1	Grid differential equations	15
4.2	Stochastic grid equations and the Fokker-Plank equation	17
5	Conclusion	23

1 Introduction

Suppose that x is a physical quantity whose evolution is governed by a deterministic force which has small random fluctuations; such a phenomenon can be described by the following equation

$$\dot{x} = f(x) + h(x)\xi(t) \quad (1)$$

*Dipartimento di Matematica Applicata "U.Dini" Via F. Buonarroti 1/c , I - 56127 Pisa
benci@dma.unipi.it, galatolo@dm.unipi.it, ghimenti@dm.unipi.it

where $\dot{x} = \frac{dx}{dt}$, and ξ is a "white noise". Intuitively, a *white noise* is the derivative of a *Brownian motion*, namely a continuous function which is not differentiable in any point.

There is no function ξ which has such a property, actually the mathematical object which models ξ is a distribution. Thus equation (1) makes sense if it lives in the world of distributions.

On the other hand the kind of problems which an applied mathematician asks are of the following type. Suppose that $x(0) = 0$ and that $\xi(t)$ is a random noise of which only the statistical properties are known. What is the probability distribution $P(t, x)$ of x at the time t ?

This question can be formalized by the theory of stochastic differential equation and eq. (1) takes the form

$$dx = f(x)dt + h(x)dw. \quad (2)$$

thus, the white noise dw is regarded as the "differential" of a Wiener process w . In this case, both $x(t)$ and $w(t)$ are modelled, not by distributions, but by stochastic processes.

By the Ito theory, the above question can be solved rigorously: the probability distribution can be determined solving the Fokker-Plank equation:

$$\frac{\partial P}{\partial t} = \frac{1}{2} \frac{\partial^2}{\partial x^2} (h(x)^2 P) + \frac{\partial}{\partial x} (f(x) P) \quad (3)$$

Eq. (1) (or (2)) and (3) are very relevant in applications of Mathematics and the practitioners of mathematics such as engineers, physicists, economists, etc. make a large use of it. However the mathematics used in these equations is rather involved and many of them are not able to control it.

Usually people think of some intuitively simpler model. For example, $\xi(t)$ is considered as a force which acts at discrete instants of time t_i ; it is supposed that the difference of two successive times $dt = t_{i+1} - t_i$ is infinitesimal and its strength is infinite; namely

$$\xi(t) = \pm \frac{1}{\sqrt{dt}} \quad (4)$$

The sign of this force is determined by a fair coin tossing. Clearly eq. (4) makes no sense and the gap between the rigorous mathematical description and the intuitive model is quite large.

The main purpose of this paper is to reduce this gap. We will use infinite and infinitesimal numbers in such a way that eqs. (4) and (1) make sense and, in this framework, we will deduce eq.(3) rigorously. Our proof is relatively simple and very close to intuition.

The use of infinite and infinitesimal numbers naturally leads to Nonstandard Analysis (*NSA*). Actually, some mathematicians have described the stochastic differential equation by Nonstandard Analysis (cf. e.g. [2], [8], [1], [11] and references therein). However the machinery of *NSA* is too complicate for practitioners of mathematics even if its ideas are simpler. A good knowledge of formal logic is necessary to follow a *NSA* proof, in fact, the main tool is the

transfer principle which, in order to be applied correctly, needs the notions of *formal language* and *interpretation*.

In this paper we will not use *NSA* but α -theory which is a kind of simplified version of Nonstandard Analysis. α -theory has been introduced in [5] (see also [4] and [3]) with the purpose to provide a simpler approach to *NSA*. In fact, in the quoted paper it has been proved that a particular model of *NSA* can be deduced by the axioms of α -theory (we refer also to [6] and to [7] for the reader interested to investigate in these questions).

The main differences between α -theory and the usual Nonstandard Analysis are two:

- α -theory does not need the language (and the knowledge) of symbolic logic;
- it does not need to distinguish two mathematical universes, (the standard universe and the nonstandard one).

α -theory postulates the "existence" of an infinite integer number called α and it provides the rules necessary to deal with the mathematical objects which can be constructed by its introduction. For example, α -theory allows to define functions such as " $\sin(\alpha t)$ " and to manage with it. α -theory is not as powerful as *NSA*, but it is simpler and it allows to treat many problems by an elementary and rigorous use of infinite and infinitesimal numbers.

In particular, using this theory, it is possible to define the "grid functions" which are functions defined for times t_i belonging to a set \mathbb{H} which models the axis of time. Using the notion of grid function, we are able to give a sense to (1) and (4) and to deduce eq. (3) rigorously.

Our approach presents the following peculiarities:

- we will rewrite eq. (1) as a "grid" differential equation:

$$\frac{\Delta x}{\Delta t} = f(x) + \xi(t) \quad (5)$$

where $\frac{\Delta x}{\Delta t}$ denotes the grid derivative (see Def. 18). From this equation, it is easy to recover both a distribution equation and a stochastic equation, and, at the same time, eq. (5) has a very intuitive meaning.

- when eq. (5) is considered from the stochastic point of view, the noise ξ is regarded as a *grid function* belonging to the space of all possible noises \mathcal{R} . If ξ is regarded as a random variable, the probability on the sample space \mathcal{R} can be defined in a naive way, namely every noise has the same probability. This is the basic idea of the Loeb measure ([9]) which is an important tool in the applications of *NSA*, but we do not need to use it. Actually we do not need to introduce any kind of measure.

2 The Alpha-Calculus

2.1 Basic notions of Alpha-Theory

In this section, we will expose the basic facts of α -theory and the basic tools which will be used in the paper in a elementary and self contained way.

α -theory is based on the existence of a new mathematical object, namely α which is added to the other entities of the mathematical universe. We may think of α as a new “*ideal*” natural number added to \mathbb{N} , in a similar way as the imaginary unit i can be seen as a new ideal number added to the real numbers \mathbb{R} . Before going to the axioms for α , we remark that *all* usual principles of mathematics are implicitly assumed. Informally, we can say that, by adopting α -theory, one can construct sets and functions according to the “usual” practice of mathematics, with no restrictions whatsoever. A precise definition of what we mean by “usual principles of mathematics” (i.e. of our underlying set theory) is given in [5].

Like the use of the imaginary entity i is governed by simple properties like $i^2 = -1$ and the usual rules for the product and sum, the use of α is governed by the following five axioms.

$\alpha 1$. Extension Axiom.

Every sequence φ can be uniquely extended to $\mathbb{N} \cup \{\alpha\}$. The corresponding value at α will be denoted by $\varphi(\alpha)$ and called the value of φ at the point α or more simply “ α -value”. If two sequences φ, ψ are different at all points, then $\varphi(\alpha) \neq \psi(\alpha)$.

We remark that if $\varphi : \mathbb{N} \rightarrow A$, then in general $\varphi(\alpha) \notin A$. The “difference preserving” condition given above can be rephrased as follows: “If two sequences are different at all n then they must be different at the point “ α ” as well”. It is a non-triviality condition, that will allow plenty of values at infinity. Moreover remark that the α -value of a sequence should not be confused with its limit. In fact, the α -value differs from a limit even for this first axiom; in fact different sequences might have the same limit.

The next axiom gives a natural coherence property with respect to compositions. If $g : A \rightarrow B$ and $h : B \rightarrow C$, denote by $h \circ g : A \rightarrow C$ the composition of h and g , i.e. $(h \circ g)(x) = h(g(x))$.

$\alpha 2$. Composition Axiom.

If φ and ψ are sequences and if f is any function such that compositions $f \circ \varphi$ and $f \circ \psi$ make sense, then

$$\varphi(\alpha) = \psi(\alpha) \Rightarrow (f \circ \varphi)(\alpha) = (f \circ \psi)(\alpha)$$

So, if two sequences takes the same value at infinity, by composing them with any function we again get sequences with the same α -value.

$\alpha 3$. Real Number Axiom.

If $c_r : n \mapsto r$ is the constant sequence with value r , then $c_r(\alpha) = r$; if $1_{\mathbb{N}} : n \mapsto n$ is the immersion of \mathbb{N} in \mathbb{R} , then $1_{\mathbb{N}}(\alpha) = \alpha \notin \mathbb{R}$.

This axiom simply says that, for real numbers, the notions of constant sequence is preserved at infinity. The latter condition says that the ideal number α is actually a *new* number. Thus the immersion $1_{\mathbb{N}}$ provides a first example of sequence $\varphi : \mathbb{N} \rightarrow \mathbb{N}$ such that $\varphi(\alpha) \notin \mathbb{N}$.

 $\alpha 4$. Internal Set Axiom.

If ψ is a sequence of sets, then also $\psi(\alpha)$ is a set and

$$\psi(\alpha) = \{\varphi(\alpha) : \varphi(n) \in \psi(n) \text{ for all } n\}.$$

Thus, the membership relation is preserved at infinity. That is, if $\varphi(n) \in \psi(n)$ for all n , then $\varphi(\alpha) \in \psi(\alpha)$. Besides, all elements of $\psi(\alpha)$ are obtained in this way. That is, they all are values at infinity of sequences which are pointwise members of ψ . The set considered above will be called *Internal sets*.

 $\alpha 5$. Pair Axiom.

If $\vartheta(n) = \{\varphi(n), \psi(n)\}$ for all n , then $\vartheta(\alpha) = \{\varphi(\alpha), \psi(\alpha)\}$.

Thus, if the sequence ξ is such that either $\xi(n) = \varphi(n)$ or $\xi(n) = \psi(n)$ for all n , then either $\xi(\alpha) = \varphi(\alpha)$ or $\xi(\alpha) = \psi(\alpha)$ at infinity as well. As a straight consequence of the last two axioms, any constant sequence with value a finite set of natural numbers, or a finite set of finite sets of natural numbers etc., takes the same value at infinity as well. We remark that this is not true in general.

We remark that the above five axioms are given somewhat “informally”. Precise indications for a rigorous formulation as sentences of a suitable first-order language are given in [5]. Also, we refer to [5] for the proves of the propositions below, but we suggest the reader to try them by himself to get acquainted with α -theory.

Definition 1 If A is a set, the $*$ -transform of A is defined as follows:

$$A^* = \{\varphi(\alpha) : \varphi : \mathbb{N} \rightarrow A\}.$$

If ψ is a sequence such that $\psi(n) = A$ for all n , then by the Internal Set Axiom, we have that $\psi(\alpha) = A^*$. Then constant set-valued sequences behave differently than real valued sequences (cf. the Real Number Axiom).

Definition 2 The set of the hyperreal numbers is the $*$ -transform of the set of the real numbers:

$$\mathbb{R}^* = \{\varphi(\alpha) : \varphi : \mathbb{N} \rightarrow \mathbb{R}\}.$$

In other words, the hyperreal numbers are the α -values assumed by real sequences. With obvious notation, for instance we will write $\sin \frac{2}{\alpha}$ to mean the hyperreal number obtained as the values at infinity of the sequence $\{\sin \frac{2}{n}\}_{n \in \mathbb{N}}$.

The sum and product operation are naturally transported on the hyperreal set moreover we have the following:

Proposition 3 *The hyperreal number system $\langle \mathbb{R}^*; +, \cdot, 0, 1, < \rangle$ is an ordered field.*

Besides the considered sets of hyper-numbers, another fundamental notion in nonstandard analysis is the following.

Definition 4 *A set $\Gamma \subset A^*$ is called hyperfinite if*

$$\Gamma = \{\varphi(\alpha) : \varphi(n) \in A_n\}$$

where $A_n \subset A$ is a sequence of finite sets. Given a hyperfinite set Γ , we define its cardinality $|\Gamma|$ as follows:

$$|\Gamma| = \psi(\alpha) \in \mathbb{N}^*$$

where $\psi(n) = |A_n|$ is the cardinality of the finite set A_n .

In general hyperfinite sets are infinite; their importance relies in the fact that they retain all “elementary” properties of finite sets. Applications of hyperfinite sets will be given in subsequent sections, for example the following hold

Proposition 5 *Every nonempty hyperfinite subset of \mathbb{R}^* has a greatest and a smallest element.*

A very important example of hyperfinite set which we will use in this paper is the *hyperfinite grid* \mathbb{H} . The *hyperfinite grid* \mathbb{H}_α is defined as the α -value of the set

$$\mathbb{H}_n = \left\{ \frac{k}{n} : k \in \mathbb{Z}, -\frac{n^2}{2} \leq k < \frac{n^2}{2} \right\};$$

namely,

$$\mathbb{H}_\alpha := \left\{ \frac{k}{\alpha} : k \in \mathbb{Z}^*, -\frac{\alpha^2}{2} \leq k < \frac{\alpha^2}{2} \right\}$$

In the following, for short, usually we will write \mathbb{H} instead of \mathbb{H}_α . Clearly \mathbb{H} is an hyperfinite set with $|\mathbb{H}| = \alpha^2$. Given $a, b \in \mathbb{H}$, we set

$$\begin{aligned} [a, b]_{\mathbb{H}} &= \{x \in \mathbb{H} : a \leq x \leq b\} \\ [a, b)_{\mathbb{H}} &= \{x \in \mathbb{H} : a \leq x < b\} \end{aligned}$$

If we identify the functions with their graphs, f^* is defined by definition 1 and it is not difficult to prove the following

Proposition 6 *Let $f : A \rightarrow B$ be a function. Then its star-transform f^* is a function $f^* : A^* \rightarrow B^*$ and, for every sequence $\varphi : \mathbb{N} \rightarrow A$,*

$$f^*(\varphi(\alpha)) = (f \circ \varphi)^*(\alpha)$$

Moreover, f^ is 1-1 (or onto) iff f is 1-1 (or onto, respectively).*

When confusion is unlikely, we will omit the symbol "*" and " f^* " will be denoted by " f ".

Let $f_n : A \rightarrow B$ be a sequence of functions; then identifying the functions with their graphs f_α is well defined by axiom ($\alpha 3$) and we have that

$$f_\alpha : A^* \rightarrow B^*$$

is a function defined by

$$f_\alpha(\varphi(\alpha)) = \psi(\alpha)$$

where $\psi(n) := f_n(\varphi(n))$ is a sequence in B .

Definition 7 *A function*

$$f : A^* \rightarrow B^*$$

is called internal if it is the graph of an internal set, namely if there is a sequence of functions $f_n : A \rightarrow B$ such that

$$f = f_\alpha$$

2.2 Infinitesimally small and infinitely large numbers.

A fundamental feature of α -calculus is that the intuitive notions of "infinitesimally small" number and "infinitely large" number can be formalized as actual objects of the hyperreal line. This give many possibilities to simplify proofs and statements in calculus theory.

Definition 8 *A hyperreal number $\xi \in \mathbb{R}^*$ is bounded or finite if its absolute value $|\xi| < r$ for some $r \in \mathbb{R}$. We say that ξ is unbounded or infinite if it is not bounded. ξ is infinitesimal if $|\xi| < r$ for all positive $r \in \mathbb{R}$.*

Clearly, the inverse of an infinite number is infinitesimal and vice versa, i.e. the inverse of a (nonzero) infinitesimal number is infinite. An example of an infinitesimal is given by $\odot := 1/\alpha$, the α -value of the sequence $\{1/n\}$.

From now on, the symbol \odot will always denote $1/\alpha$.

All infinitesimal and all real numbers are bounded. However there are finite hyperreals that are neither infinitesimal nor real, for example $5 + \odot$ and $7 + \sin \alpha$.

Definition 9 *We say that two hyperreal numbers ξ and η are infinitesimally close if $\xi - \eta$ is infinitesimal. In this case we write $\xi \sim \eta$.*

It is easily seen that \sim is an equivalence relation.

On the other hand (as it is intuitive) each bounded hyperreal is infinitely close to some real. The following indeed comes from the completeness of the real line.

Theorem 10 (Shadow Theorem) *Every bounded hyperreal number ξ is infinitesimally close to a unique real number r , called the shadow of ξ . Symbolically $r = sh(\xi)$.*

The notion of a shadow is extended to every hyperreal number, by setting $sh(\xi) = +\infty$ if ξ is positive unbounded, and $sh(\xi) = -\infty$ if ξ is negative unbounded.

Definition 11 *Given two hyperreal numbers ξ and $\zeta \in \mathbb{R}^* \setminus \{0\}$, we say that they have the same order if ξ/ζ and ζ/ξ are bounded numbers and we will write*

$$\xi \approx \zeta$$

(notice the difference between " \sim " and " \approx " since these symbols will be largely used in the rest of this paper). We say that ξ has a larger order than ζ if ξ/ζ is an infinite number and we will write

$$\xi \gg \zeta$$

We say that ξ has a smaller order than ζ if ξ/ζ is an infinitesimal number and we will write

$$\xi \ll \zeta$$

2.3 Some notions of infinitesimal calculus

Now we see how all this machinery can be used to build a rigorous "infinitesimal" calculus. We present how the definition of limit can be given in our setting.

Definition 12 *We say that $\lim_{x \rightarrow x_0} f(x) = l$ if $f^*(\xi) \sim l$ for all $\xi \sim x_0$ ($\xi \neq x_0$).*

With the definition of limit all the elementary calculus can be reconstructed, but the features of our method allow to avoid the use of limits and work with *real* infinitesimal and infinite numbers. Let us see some example: the definition of continuity and derivative. We remark that the theory given by these definitions is equivalent to the standard calculus and all the known results (as for example the Lagrange's or Fermat's theorems) applies.

Definition 13 *A real function $f : A \rightarrow \mathbb{R}$ is continuous at $x_0 \in A$ if for every $\xi \in A^*$, $\xi \sim x_0 \Rightarrow f^*(\xi) \sim f^*(x_0)$.*

Let f be any real function defined on a neighborhood of x_0 .

Definition 14 We say that f has derivative at x_0 if there exists $f'(x_0) \in \mathbb{R}$ such that for all infinitesimals $\varepsilon \neq 0$,

$$\frac{f^*(x_0 + \varepsilon) - f^*(x_0)}{\varepsilon} \sim f'(x_0)$$

Equivalently, f has derivative $f'(x_0)$ at x_0 if for every infinitesimal ε there is an infinitesimal δ such that $f^*(x_0 + \varepsilon) = f(x_0) + f'(x_0)\varepsilon + \delta\varepsilon$.

As said before all the classical results of calculus hold in this framework. An example which will be used in the following is the Taylor formula (with infinitesimal remainder).

Theorem 15 If $f \in C^{n+1}(\mathbb{R})$ then for each infinitesimal ϵ there is an infinitesimal η such that

$$f^*(x + \epsilon) = \sum_{k \leq n} \frac{f^{(k)}(x)\epsilon^k}{k!} + \eta\epsilon^n.$$

Now we introduce a concept of integral. This concept is more general than the Riemann integral and will allow us to integrate noises and stochastic equations. Intuitively this integral is just an infinite sum of hyperreal numbers. This sum will be done on an hyperfinite set.

Definition 16 If $\Gamma = \chi(\alpha)$ is a hyperfinite set of hyperreal numbers, then its hyperfinite sum:

$$\sum_{x \in \Gamma} x = \text{Sum}_\chi(\alpha)$$

is defined as the value at infinity of the sequence of finite sums

$$\text{Sum}_\chi(n) = \sum_{x \in \chi(n)} x.$$

It is easily checked that this definition does not depend on the choice of the sequence $\{\chi(n)\}$, but only on its value at infinity Γ . Using this definition, we define the α -integral.

Definition 17 Let $f : A \rightarrow \mathbb{R}$ be any function, where $A \subseteq \mathbb{R}$. Its Alpha-integral on A , denoted by $\int_A f(x) \Delta x$, is the number in $\mathbb{R} \cup \{\pm\infty\}$ defined as the shadow of the following hyperfinite sum:

$$\int_A f(x) \Delta x = sh \left(\oplus \cdot \sum_{\xi \in \mathbb{H} \cap A^*} f^*(\xi) \right)$$

Notice that

$$\int_A f(x) \Delta x = sh(S_A(\alpha)) \text{ where } S_A(n) = \frac{1}{n} \cdot \sum_{x \in \mathbb{H}(n) \cap A} f^*(x)$$

Of course, if $A = [a, b]$ is a closed interval, we adopt the usual notation $\int_a^b f(x) \Delta x$.

The Alpha-integral $\int_a^b f(x) \Delta x$ is defined for every function. In fact, while the sequence

$$S_a^b(n) = \frac{1}{n} \cdot \sum_{x \in \mathbb{H}(n) \cap (a, b)} f^*(x)$$

may not have a limit in the classic sense, its α -value $S_a^b(\alpha)$ is always defined. If the function f is Riemann integrable then $\lim_{n \rightarrow \infty} S_a^b(n)$ exists and coincides with the α -integral (notice that if a real sequence $\{\varphi(n)\}$ has “classic” limit $l \in \mathbb{R} \cup \{\pm\infty\}$, then it must be $sh(\varphi(\alpha)) = l$). Thus the Alpha-integral actually generalizes Riemann integral.

3 Grid functions

A grid function is a function whose argument range on an hyperfinite “grid” whose elements are the (hypernatural) multiples of $\frac{1}{\alpha}$. Since the grid is hyperfinite these functions are easy to handle and from many points of view they behave similarly to functions on finite sets. We will see that this simple kind of functions are flexible enough to contain elements representing distributions. This flexibility will allow us to obtain in a simple way a kind of stochastic calculus (see, e.g. the Ito’s formula, Thm. 21).

3.1 Basic notions

An internal function

$$\xi : \mathbb{H} \rightarrow \mathbb{R}^*$$

is called *grid function*.

Definition 18 Given a grid function $\xi : \mathbb{H} \rightarrow \mathbb{R}^*$, we define its grid derivative $\frac{\Delta \xi}{\Delta t}$ as

$$\frac{\Delta \xi}{\Delta t}(t) = \frac{\xi(t + \ominus) - \xi(t)}{\ominus};$$

The grid integral of ξ is defined as

$$\mathbb{I}[\xi] = \ominus \sum_{t \in \mathbb{H}} \xi(t);$$

if $\Gamma \subset \mathbb{H}$ is a hyperfinite set we define $\mathbb{I}_\Gamma[\xi]$, its grid integral in Γ , as

$$\mathbb{I}_\Gamma[\xi] = \ominus \sum_{t \in \Gamma} \xi(t)$$

Most of the properties of the usual derivative hold also for the grid derivative, for example we have that, if ξ and ζ are continuous functions, with finite grid derivative,

$$\begin{aligned}\frac{\Delta(\xi\zeta)}{\Delta t} &= \frac{\xi(t + \ominus)\zeta(t + \ominus) - \xi(t)\zeta(t)}{\ominus} = \\ &= \frac{\xi(t + \ominus)\zeta(t + \ominus) - \xi(t + \ominus)\zeta(t) + \xi(t + \ominus)\zeta(t) - \xi(t)\zeta(t)}{\ominus} = \\ &= \frac{\Delta\xi}{\Delta t}(t) \cdot \zeta(t) + \xi(t + \ominus) \cdot \frac{\Delta\zeta}{\Delta t}(t) \sim \frac{\Delta\xi}{\Delta t} \cdot \zeta + \xi \cdot \frac{\Delta\zeta}{\Delta t}.\end{aligned}$$

These notions can be easily extended to functions of more variables; for example if

$$\rho(t, x) : \mathbb{H} \times \mathbb{H} \rightarrow \mathbb{R}^*$$

we set

$$\begin{aligned}\frac{\Delta\rho}{\Delta t}(t, x) &= \frac{\rho(t + \ominus, x) - \rho(t, x)}{\ominus} \\ \frac{\Delta\rho}{\Delta x}(t, x) &= \frac{\rho(t, x + \ominus) - \rho(t, x)}{\ominus}\end{aligned}$$

and if $\Gamma \subset \mathbb{H}^2$ is a hyperfinite set we define its grid integral $\mathbb{I}_\Gamma[\rho]$ as

$$\mathbb{I}_\Gamma[\rho] = \ominus^2 \sum_{(t, x) \in \Gamma} \rho(t, x).$$

It is clear that the derivative of a grid function ξ is a grid function. Moreover, if ξ is a grid function, then the *grid integral function* $x \mapsto \mathbb{I}_{[a, x]}[\xi]$ is a grid function. We have the following relation between the grid-derivative and the grid-integral:

Theorem 19 *If ξ is a grid function, then*

$$\begin{aligned}\mathbb{I}_{[x, y]} \left[\frac{\Delta\xi}{\Delta x} \right] &= \xi(y) - \xi(x) \\ \frac{\Delta}{\Delta x} \mathbb{I}_{[a, x]}[\xi] &= \xi(x)\end{aligned}$$

Proof. *Obviously we have*

$$\begin{aligned}\mathbb{I}_{[x, y]} \left[\frac{\Delta\xi}{\Delta x} \right] &= \ominus \sum \frac{\xi(x + \ominus) - \xi(x)}{\ominus} + \frac{\xi(x + 2 \cdot \ominus) - \xi(x + \ominus)}{\ominus} \dots + \frac{\xi(y) - \xi(y - \ominus)}{\ominus} = \\ &= \xi(y) - \xi(x).\end{aligned}$$

Furthermore

$$\begin{aligned}\frac{\Delta}{\Delta x} \mathbb{I}_{[a, x]}[\xi] &= \frac{\mathbb{I}_{[a, x + \ominus]}[\xi] - \mathbb{I}_{[a, x + \ominus]}[\xi]}{\ominus} = \\ &= \sum_{t \in [a, x + \ominus)} \xi(t) - \sum_{t \in [a, x)} \xi(t) = \xi(x).\end{aligned}$$

■

Definition 20 of A grid function ξ is called integrable in $[a, b]$ if $\mathbb{I}_{[a, b]} [\xi]$ is finite; in this case, we set

$$\int_a^b \xi(s) \Delta s := sh \left(\mathbb{I}_{[a, b]} [\xi] \right) = sh \left(\bigoplus_{t \in \mathbb{H} \cap [a, b)} \xi(t) \right)$$

ξ is called absolutely integrable in $[a, b]$ if $\mathbb{I}_{[a, b]} [|\xi|]$ is finite. $\int_a^b \xi(s) ds$ will be called α -integral of ξ .

Of course, this integral is strictly related to the α -integral given in Def. 17. In fact, to every real function

$$f : [a, b] \rightarrow \mathbb{R}$$

it is possible to associate its natural extension

$$f^* : [a, b]^* \rightarrow \mathbb{R}^*$$

and a grid function

$$\tilde{f} : [a, b]_{\mathbb{H}} \rightarrow \mathbb{R}^* \quad (6)$$

obtained as restriction of f^* to $[a, b]_{\mathbb{H}}$. When no ambiguity is possible we will denote f^* and \tilde{f} with the same symbol.

The α -integral of f coincides with the α -integral of \tilde{f} given by Def. 20.

3.2 The Ito formula

We show the power of the grid functions approach by stating in a very simple way a proposition which is, in some sense, a variant of the Ito's formula. As in the standard approach this formula will be the main tool in the study of grid stochastic equations.

Theorem 21 (Nonstandard Ito's Formula) Let $\varphi \in C_0^3(\mathbb{R}^2)$ and $x(t)$ be a grid function such that

$$\left| \frac{\Delta x}{\Delta t}(t) \right| \leq \eta \alpha^{2/3}, \quad (7)$$

where $\eta \sim 0$.

Then

$$\frac{\Delta}{\Delta t} \varphi(t, x(t)) \sim \varphi_t(t, x(t)) + \varphi_x(t, x(t)) \frac{\Delta x}{\Delta t}(t) + \frac{\bigoplus}{2} \varphi_{xx}(t, x(t)) \cdot \left(\frac{\Delta x}{\Delta t}(t) \right)^2.$$

Here φ_t , φ_x and φ_{xx} denote the usual partial derivative of φ .

Proof. By definition of grid derivative we have that

$$\begin{aligned}
\frac{\Delta}{\Delta t} \varphi(t, x(t)) &= \frac{\varphi(t + \mathbb{E}, x(t + \mathbb{E})) - \varphi(t, x(t + \mathbb{E}))}{\mathbb{E}} + \frac{\varphi(t, x(t + \mathbb{E})) - \varphi(t, x(t))}{\mathbb{E}} \\
&\sim \varphi_t(t, x(t + \mathbb{E})) + \frac{\varphi(t, x(t + \mathbb{E})) - \varphi(t, x(t))}{\mathbb{E}} \\
&\sim \varphi_t(t, x(t)) + \frac{\varphi(t, x(t + \mathbb{E})) - \varphi(t, x(t))}{\mathbb{E}}
\end{aligned}$$

But

$$\varphi(t, x(t + \mathbb{E})) = \varphi\left(t, x(t) + \mathbb{E} \frac{\Delta x}{\Delta t}(t)\right),$$

and $|\mathbb{E} \frac{\Delta x}{\Delta t}(t)| \leq \eta \alpha^{2/3} \cdot \mathbb{E} = \eta \cdot \mathbb{E}^{1/3}$ is infinitesimal. Then, using the Taylor formula (Theorem 15), we have that

$$\begin{aligned}
\varphi\left(t, x(t) + \mathbb{E} \frac{\Delta x}{\Delta t}(t)\right) &= \varphi(t, x(t)) + \varphi_x(t, x(t)) \cdot \mathbb{E} \frac{\Delta x}{\Delta t}(t) \\
&\quad + \frac{1}{2} \varphi_{xx}(t, x(t)) \left(\mathbb{E} \frac{\Delta x}{\Delta t}(t)\right)^2 \\
&\quad + \frac{1}{3!} \varphi_{xxx}(t, x(t)) \left(\mathbb{E} \frac{\Delta x}{\Delta t}(t)\right)^3 + \delta \left(\mathbb{E} \frac{\Delta x}{\Delta t}(t)\right)^3
\end{aligned}$$

where δ is an infinitesimal; hence

$$\begin{aligned}
\frac{\varphi(t, x(t + \mathbb{E})) - \varphi(t, x(t))}{\mathbb{E}} &= \varphi_x(t, x(t)) \frac{\Delta x}{\Delta t}(t) + \frac{\mathbb{E}}{2} \varphi_{xx}(t, x(t)) \cdot \left(\frac{\Delta x}{\Delta t}(t)\right)^2 \\
&\quad + \frac{\mathbb{E}^2}{6} \varphi_{xxx} \cdot \left(\frac{\Delta x}{\Delta t}(t)\right)^3 + \delta \cdot \mathbb{E}^2 \left(\frac{\Delta x}{\Delta t}(t)\right)^3
\end{aligned}$$

By the assumption (7) the last two terms are infinitesimal and we get the required result. ■

3.3 Distributions and grid functions

The grid functions can be considered as a sort of generalization of the usual real functions.

In fact to every real function correspond a unique grid functions given by (6). In the traditional analysis the most important generalization of the real function is given by the distribution. In this section we will show that the grid functions represent also a generalization of the notion of *distribution*.

First of all we recall some notation: given a set $A \subset \mathbb{R}^N$, $\mathcal{D}(A)$ denotes the space of C^∞ functions with compact support of A . The space of the distributions $\mathcal{D}'(A)$ is the topological dual of $\mathcal{D}(A)$ when $\mathcal{D}(A)$ is equipped with the Schwartz topology.

Actually, $\mathcal{D}'(A)$ can also be constructed without knowing the Schwartz topology by using the notion of grid function. Next, we will show how to do it.

Let $\mathfrak{G}(A)$ denote the set of grid function defined on

$$A_{\mathbb{H}} := A^* \cap \mathbb{H}^N$$

On $\mathfrak{G}(A)$ we define the following equivalence relation:

Definition 22 *Two grid functions ξ_1, ξ_2 are said to be equivalent if*

$$\forall \varphi \in \mathcal{D}, \quad \int (\xi_1 - \xi_2) \varphi ds = 0$$

In this case we will write

$$\xi_1 \sim_{\mathcal{D}} \xi_2$$

We may think that two grid functions are equivalent if they are *macroscopically equal*.

Moreover, we set

$$\mathfrak{G}_0(A) = \{\xi \in \mathfrak{G}(A) : \forall \varphi \in \mathcal{D}, \mathbb{I}_{A_{\mathbb{H}}}[\xi \varphi] \text{ is finite}\}$$

The set of distributions $\mathcal{D}'(A)$ can be defined as follows

$$\mathcal{D}'(A) = \frac{\mathfrak{G}_0(A)}{\sim_{\mathcal{D}}}.$$

Thus a distribution can be considered as an equivalence class T_{ξ} of some grid function $\xi \in \mathfrak{G}_0(A)$.

T_{ξ} can be identified with an element of $\mathcal{D}'(A)$ by the following formula:

$$\langle T_{\xi}, \varphi \rangle = \int_A \xi \varphi ds = sh \left(\bigodot \cdot \sum_{t \in A_{\mathbb{H}}} \xi(t) \varphi^*(t) \right), \quad \varphi \in \mathcal{D}. \quad (8)$$

To each distribution we can associate a grid function. For example, if $T \in \mathcal{D}'(\mathbb{R})$ we can do in the following way. Since a distribution T has the following representation¹:

$$T = \sum_{k=0}^{\infty} D^k f_k$$

where f_k are continuous. Then the grid function ξ corresponding to T is given by

$$\xi(t) = \sum_{k=0}^{\alpha} \frac{\Delta^k}{\Delta t^k} f_k(t).$$

Let us see some simple example. The function

$$\delta(t) = \alpha \delta_{0,t}$$

¹See Rudin, functional analysis, Th. 6.28, pag.169

where $\delta_{i,j}$ is the Kronecker symbol correspond to the Dirac δ . But also the following grid functions

$$\alpha \frac{\delta_{0,t} + \delta_{\Delta,t}}{2}; \sum \delta_{0,t+\frac{k}{\alpha}} \quad (k \in \mathbf{Z}); \text{ etc}$$

correspond to the Dirac δ . The grid function

$$\frac{\Delta \delta}{\Delta t}(t) = \alpha^2 (\delta_{0,t} - \delta_{\ominus,t})$$

correspond to δ' .

The grid function $\alpha^2 \delta_{0,t}$ is not in $\mathfrak{G}_0(A)$ and hence it does not correspond to any distribution.

4 Stochastic differential equations

4.1 Grid differential equations

A grid ordinary differential equation is a differential equation whose time step ranges on the hyperfinite grid. This fact makes it to work as a discrete time object simplifying many formal aspects.

A grid ordinary differential equation is then an equation of the kind

$$\frac{\Delta x}{\Delta t}(t) = f(t, x(t)), \quad (9)$$

where $t \in \mathbb{H}$, $x(t)$ is a grid function and $f : \mathbb{H} \times \mathbb{R}^* \rightarrow \mathbb{R}^*$ is an internal function. A grid function $x(t)$ is a solution of the grid equation if satisfies it at each point of the grid.

The following result shows that such an equation has an unique solution. This, without regularity assumptions on the f . Hence, this kind of equations has solutions even if the equations contain a noise term (see section (4.2)).

Theorem 23 *Given an initial time $t_0 \in \mathbb{H}$ and an initial data $x_0 \in \mathbb{R}^*$, the Cauchy problem associated to (9), that is*

$$\begin{cases} \frac{\Delta x}{\Delta t}(t) = f(t, x(t)) & t \in \mathbb{H} \\ x(t_0) = x_0 \end{cases} \quad (10)$$

admits for $t \geq t_0$ an unique solution $x : \mathbb{H} \rightarrow \mathbb{R}^$.*

Proof. We know that f is an ideal value of a sequence $\{f_n\}_{n \in \mathbb{N}}$. Also, we have that $t_0 = t_{0,\alpha}$, $x_0 = x_{0,\alpha}$ are the ideal values associated to $\{t_{0,n}\}_{n \in \mathbb{N}}$, $\{x_{0,n}\}_{n \in \mathbb{N}}$. For each $n \in \mathbb{N}$ and $m \in \mathbb{Z}$, we can construct by induction a sequence of functions.

$$x_n : \frac{1}{n}\mathbb{Z} \rightarrow \mathbb{R}$$

as follows:

$$x_n(t_{0,n}) = x_{0,n} \quad (11)$$

$$x_n\left(t_{0,n} + \frac{m+1}{n}\right) = x_n\left(t_{0,n} + \frac{m}{n}\right) + \quad (12)$$

$$+ \frac{1}{n} f_n\left(\left(t_{0,n} + \frac{m}{n}\right), x_n\left(t_{0,n} + \frac{m}{n}\right)\right). \quad (13)$$

Then by definition of internal function we have that, for $t = t_0 + m \cdot \ominus$, $m \in \mathbb{Z}^*$

$$x_\alpha(t_0) = x_0 \quad (14)$$

$$x_\alpha(t + \ominus) = x_\alpha(t) + \ominus f_\alpha(t, x_\alpha(t)). \quad (15)$$

Thus $x = x_\alpha$ solves (10). It is easy to check that this solution is also unique.

■

Given $x_0 \sim x_1$, it may happen that

$$x(t, x_0) \approx x(t, x_1)$$

where $x(t, x_i)$ is the solution of (10) with initial data x_i . Some times we would like to have

$$x(t, x_0) \sim x(t, x_1) \quad \forall x_1 \sim x_0; \quad (16)$$

this can be useful, for example, when we want to consider the standard part of a hyperreal differential equation.

We have the following proposition :

Proposition 24 *Consider the following Cauchy problem*

$$\begin{cases} \frac{\Delta x}{\Delta t}(t) = f(t, x(t)) & t \in \mathbb{H}; \\ x(t_0) = x_0, \end{cases} \quad (17)$$

and suppose that, there exists L s.t.

$$|f(t, x) - f(t, y)| \leq L|x - y|. \quad (18)$$

Let x_1 be a bounded initial data for the problem (17) and $x(t, x_1)$ is the solution of this problem. Then , if $x_1 \sim x_0$, then for all $0 \leq t < T$ we have

$$x(t, x_0) \sim x(t, x_1) \quad (19)$$

Proof. Arguing as in standard analysis, we have that (18) guarantees that, for any $T_1 < T$, the solution is bounded. Moreover, in standard analysis, the condition (18) guarantees, the continuous dependence of the solution from initial data x_0 . In our case, until the solution is finite, we can proceed in the same way to prove that, chosen an arbitrary $T_1 < T$

$$|x(t, x_0) - x(t, x_1)| \leq |x_0 - x_1| e^{LT_1} \quad (20)$$

for all $t \in [0, T_1]$. Because $x_0 \sim x_1$ by hypothesis, we have that

$$x(t, x_0) \sim x(t, x_1) \quad (21)$$

for all $t \in [0, T_1]$. This assures the proof. ■

4.2 Stochastic grid equations and the Fokker-Plank equation

In our approach, a stochastic differential equation consists of a set of grid differential equations. Each differential equation has a *noise* term and gives a trajectory which can be considered as a realization of a process.

Let $\mathcal{R} \subset \mathfrak{G}[0, 1]$ be an hyperfinite set of grid functions and consider the class of Cauchy problems

$$\begin{cases} \frac{\Delta x}{\Delta t}(t) = f(t, x) + h(t, x)\xi, \\ x(0) = x_0, \\ \xi(t) \in \mathcal{R}. \end{cases} \quad (22)$$

where

$$f, h : [0, 1]_{\mathbb{H}} \times \mathbb{R}^* \rightarrow \mathbb{R}^*$$

We want to study the statistical behavior of the set of solutions of the above Cauchy problems

$$\mathcal{S} = \{x_{\xi}(t) : \xi \in \mathcal{R}\};$$

More precisely we want to describe the behavior of the density function

$$\rho : [0, 1]_{\mathbb{H}} \times \mathbb{H} \rightarrow \mathbb{Q}^*$$

defined as follows

$$\rho(t, x) = \frac{|\{x_{\xi} \in \mathcal{S} : x \leq x_{\xi}(t) < x + \ominus\}|}{\ominus |\mathcal{R}|}.$$

We are interested in the case in which \mathcal{R} models a *white noise*; roughly speaking we can define a white noise as the hyperfinite set of all the grid functions with values $\pm\sqrt{\alpha}$. Here there is its precise definition:

Definition 25 *The white noise is the set of grid functions defined by*

$$\mathcal{R} = \mathcal{R}_{\alpha}$$

where

$$\mathcal{R}_n = \{-\sqrt{n}, +\sqrt{n}\}^{[0, 1]_{\mathbb{H}_n}}$$

Hence \mathcal{R} is a hyperfinite set with $|\mathcal{R}| = 2^{\alpha+1}$.

Remark 26 *We would be tempted to write*

$$\mathcal{R}_{\alpha} = \{-\sqrt{\alpha}, +\sqrt{\alpha}\}^{[0, 1]_{\mathbb{H}_{\alpha}}}$$

however this notation is very ambiguous; in fact \mathcal{R}_{α} is a set defined by the Internal Set Axiom and it contains only internal function. However the symbol $\{-\sqrt{\alpha}, +\sqrt{\alpha}\}^{[0, 1]_{\mathbb{H}_{\alpha}}}$ usually represents the set of all the functions $f : [0, 1]_{\mathbb{H}_{\alpha}} \rightarrow \{-\sqrt{\alpha}, +\sqrt{\alpha}\}$.

Now, we can state the main result of this paper:

Theorem 27 *Assume that \mathcal{R} is a white noise and that $f(t, x)$ and $h(t, x)$ are continuous functions. Then the distribution T_ρ relative to the density function ρ is a measure and satisfies the Fokker-Plank equation*

$$\frac{dT_\rho}{dt} + \frac{d}{dx} (f(t, x)T_\rho) - \frac{1}{2} \frac{d^2}{dx^2} (h(t, x)^2 T_\rho) = 0. \quad (23)$$

$$T_\rho(0, x) = \delta \quad (24)$$

in the sense of distribution.

Remark 28 *We recall that (23) and (24) "in the sense of distributions" mean that T_ρ satisfies the equation*

$$\left\langle \varphi_t + f\varphi_x + \frac{1}{2}h^2\varphi_{xx}, T_\rho \right\rangle + \varphi(x_0) = 0 \quad (25)$$

for any $\varphi \in \mathcal{D}([0, 1] \times \mathbb{R})$. The duality $\langle \cdot, \cdot \rangle$ is between the space of continuous function and the space of measures. Equation (25) can be expressed using the grid function ρ and the α -integral by the following equation:

$$\forall \varphi \in \mathcal{D}([0, 1] \times \mathbb{R}), \quad \iint (\varphi_t + f\varphi_x + \varphi_{xx}h^2) \rho \Delta x \Delta t + \varphi(0, x_0) = 0 \quad (26)$$

Actually, we will prove Th. 27 just proving the above equation.

Remark 29 *If $f(t, x)$ and $h(t, x)$ are smooth functions, by standard results in PDE, we know that, for $t > 0$, the distribution T_ρ coincides with a smooth function $u(t, x)$. Then, for any $t > 0$, ρ defines a smooth function u by the formula*

$$\forall \varphi \in \mathcal{D}((0, 1) \times \mathbb{R}), \quad \iint \rho \varphi \Delta x \Delta t = \iint u \varphi dx dt$$

and u satisfies the Fokker-Plank equation in $(0, 1) \times \mathbb{R}$ in the usual sense.

Remark 30 *We will see in the proof of Th. (27) that if the functions $f(t, x)$ and $h(t, x)$ are not continuous, but only bounded on compact sets, the equation (26) still holds. However in this case, equation (26) cannot be interpreted so easily. For example, if $f(t, x)$ and $h(t, x)$ are not measurable, there is no simple standard interpretation.*

Given $t \in [0, 1]_{\mathbb{H}}$ we set

$$\mathcal{R}[0, t) = \mathcal{R}_\alpha[0, t); \quad \mathcal{R}_n := \{-\sqrt{n}, +\sqrt{n}\}^{[0, t)_{\mathbb{H}_n}}; \quad (27)$$

namely, $\mathcal{R}[0, t]$ is the set of the restrictions of the functions of \mathcal{R} to $[0, t]_{\mathbb{H}}$. Moreover, for $\tau \in \mathcal{R}[0, s]$, we set

$$\mathcal{R}_{\tau}[s, 1] = \{\xi \in \mathcal{R} : \xi(t) = \tau(t) \text{ for } t < s\}$$

So we have the following decomposition:

$$\mathcal{R} = \bigcup_{\tau \in \mathcal{R}[0, s]} \mathcal{R}_{\tau}[s, 1]. \quad (28)$$

We define the mean value of a grid function in the set $[x, y] \cap \mathbb{H}$ as follows:

$$\mathbb{E}_{[x, y]}[f] = \frac{1}{(y - x)} \mathbb{I}_{[x, y]}[f] = \frac{\ominus}{(y - x)} \sum_{t \in [x, y] \cap \mathbb{H}} f(t)$$

In general, if Γ is a hyperfinite set and $\Phi : \Gamma \rightarrow \mathbb{R}^*$ is an internal function, the mean value of Φ in Γ is defined as follows:

$$\mathbb{E}_{\xi \in \Gamma}[\Phi] = \frac{1}{|\Gamma|} \sum_{\xi \in \Gamma} \Phi(\xi)$$

Proposition 31 *If \mathcal{R} is a white noise, then for any $t \in [0, 1]_{\mathbb{H}}$, and $\tau \in \mathcal{R}[0, t]$, we have*

$$\text{the hyperfinite number } |\mathcal{R}_{\tau}[t, 1]| \text{ does not depend on } \tau \in \mathcal{R}[0, t] \quad (29)$$

and

$$\mathbb{E}_{\xi \in \mathcal{R}_{\tau}[t, 1]}[\xi(t)] \sim 0 \quad (30)$$

$$\mathbb{E}_{\xi \in \mathcal{R}_{\tau}[t, 1]}[\xi(t)^2] \sim \alpha, \quad (31)$$

Proof. The proof is almost immediate: first of all we have that

$$|\mathcal{R}_{\tau}[t, 1]| = 2^{\alpha(1-t)+1};$$

moreover

$$\begin{aligned} \mathbb{E}_{\xi \in \mathcal{R}_{\tau}[t, 1]}[\xi(t)] &= \frac{1}{|\mathcal{R}_{\tau}[t, 1]|} \sum_{\xi \in \mathcal{R}_{\tau}[t, 1]} \xi(t) \\ &= \frac{1}{2|\mathcal{R}_{\tau}[t, 1]|} \sqrt{\alpha} - \frac{1}{2|\mathcal{R}_{\tau}[t, 1]|} \sqrt{\alpha} = 0 \sim 0 \end{aligned}$$

and

$$\mathbb{E}_{\xi \in \mathcal{R}_{\tau}[t, 1]}[\xi(t)^2] = \frac{1}{|\mathcal{R}_{\tau}[t, 1]|} \sum_{\xi \in \mathcal{R}_{\tau}[t, 1]} \xi(t)^2 = \alpha$$

■

Remark 32 The conclusion of Th. 27 hold not only if the "stochastic class" \mathcal{R} is defined by (27), but for any class \mathcal{R} which satisfies the properties (29), (30) and (31). For example we can take

$$\mathcal{R} = \mathcal{R}_\alpha; \quad \mathcal{R}_n := \{q_1\sqrt{n}, \dots, q_k\sqrt{n}\}^{[0,1]_{\mathbb{H}_n}}; \quad k \in \mathbb{N}$$

with $q_i \in \mathbb{R}^*$,

$$\sum_{i=1}^k q_i = 0; \quad \sum_{i=1}^k q_i^2 = 1.$$

The following two lemmas are a direct consequence of properties (29), (30) and (31).

Lemma 33 Let $G : [0, 1]_{\mathbb{H}} \times \mathbb{R}^* \times \mathbb{R}^* \rightarrow \mathbb{R}^*$ be any internal function. Then, for every $t \in [\ominus, 1]$

$$\mathbb{E}_{\xi \in \mathcal{R}} [G(t, x_\xi(t), \xi(t))] = \mathbb{E}_{\tau \in \mathcal{R}_{[0,t]}} \mathbb{E}_{\xi \in \mathcal{R}_\tau[t,1]} [G(t, x_\xi(t), \xi(t))]$$

Proof. By (28), we have that

$$\mathcal{R} = \bigcup_{\tau \in \mathcal{R}_{[0,t]}} \mathcal{R}_\tau[t, 1]$$

Then,

$$\begin{aligned} \mathbb{E}_{\xi \in \mathcal{R}} [G(t, x_\xi(t), \xi(t))] &= \frac{1}{|\mathcal{R}_\tau[t, 1]| \cdot |\mathcal{R}_{[0,t]}|} \sum_{\tau \in \mathcal{R}_{[0,t]}} \sum_{\xi \in \mathcal{R}_\tau[t,1]} G(t, x_\xi(t), \xi(t)) \\ &= \frac{1}{|\mathcal{R}_{[0,t]}|} \sum_{\tau \in \mathcal{R}_{[0,t]}} \frac{1}{|\mathcal{R}_\tau[t, 1]|} \sum_{\xi \in \mathcal{R}_\tau[t,1]} G(t, x_\xi(t), \xi(t)) \\ &= \mathbb{E}_{\tau \in \mathcal{R}_{[0,t]}} \mathbb{E}_{\xi \in \mathcal{R}_\tau[t,1]} [G(t, x_\xi(t), \xi(t))]. \end{aligned}$$

■

Lemma 34 Let $F : [0, 1]_{\mathbb{H}} \times \mathbb{R}^* \rightarrow \mathbb{R}^*$ be an internal function such that $|F(t, x)| \leq M$, $M \in \mathbb{R}$. Then, for every $t \in [0, 1]$

$$\begin{aligned} \mathbb{E}_{\xi \in \mathcal{R}} [F(t, x_\xi(t)) \cdot \xi(t)] &\sim 0. \\ \mathbb{E}_{\xi \in \mathcal{R}} [F(t, x_\xi(t)) \cdot \xi(t)^2] &\sim \alpha \cdot \mathbb{E}_{\xi \in \mathcal{R}} [F(t, x_\xi(t))] \end{aligned}$$

Proof. By lemma 33, we have that

$$\mathbb{E}_{\xi \in \mathcal{R}} [F(t, x_\xi(t)) \cdot \xi(t)] = \mathbb{E}_{\tau \in \mathcal{R}_{[0,t]}} \mathbb{E}_{\xi \in \mathcal{R}_\tau[t,1]} [F(t, x_\xi(t)) \cdot \xi(t)]$$

Since $x_\xi(t)$ does not depend on $\xi(s)$ for $s > t$, we have that

$$\begin{aligned} \mathbb{E}_{\xi \in \mathcal{R}_\tau[t,1]} [F(t, x_\xi(t)) \cdot \xi(t)] &= \frac{1}{|\mathcal{R}_\tau[t, 1]|} \sum_{\tau \in \mathcal{R}_{[0,t]}} (F(t, x_\xi(t)) \cdot \xi(t)) \\ &= F(t, x_\xi(t)) \cdot \frac{1}{|\mathcal{R}_\tau[t, 1]|} \sum_{\tau \in \mathcal{R}_{[0,t]}} \xi(t) \\ &= F(t, x_\xi(t)) \cdot \mathbb{E}_{\xi \in \mathcal{R}_\tau[t,1]} [\xi(t)] \end{aligned}$$

Then since F is bounded, by (30), we get the conclusion:

$$\mathbb{E}_{\xi \in \mathcal{R}} [F(t, x_\xi(t)) \xi(t)] = \mathbb{E}_{\tau \in \mathcal{R}[0, t)} [F(t, x_\xi(t)) \cdot \mathbb{E}_{\xi \in \mathcal{R}_\tau[t, 1]} [\xi(t)]] \sim 0$$

Analogously, we have that

$$\mathbb{E}_{\xi \in \mathcal{R}} [F(t, x_\xi(t)) \xi(t)^2] = \mathbb{E}_{\tau \in \mathcal{R}[0, t)} [F(t, x_\xi(t)) \mathbb{E}_{\xi \in \mathcal{R}_\tau[t, 1]} [\xi(t)^2]]$$

and by (31), we get that

$$F(t, x_\xi(t)) \mathbb{E}_{\xi \in \mathcal{R}_\tau[t, 1]} [\xi(t)^2] = F(t, x_\xi(t)) (\alpha + \varepsilon_\tau)$$

where $\varepsilon_\tau \sim 0$. Then

$$\begin{aligned} \mathbb{E}_{\xi \in \mathcal{R}} [F(t, x_\xi(t)) \xi(t)^2] &= \alpha \mathbb{E}_{\tau \in \mathcal{R}[0, t)} [F(t, x_\xi(t))] + \mathbb{E}_{\tau \in \mathcal{R}[0, t)} [F(t, x_\xi(t)) \varepsilon_\tau] \\ &\sim \alpha \mathbb{E}_{\tau \in \mathcal{R}[0, t)} [F(t, x_\xi(t))] \end{aligned}$$

This concludes the proof. ■

Now we see a basic property of the density function.

Lemma 35 *Let $\varphi \in \mathcal{D}([0, 1] \times \mathbb{R})$ and let $x_\xi(t)$, $\xi \in \mathcal{R}$, be the family of solutions of a grid stochastic ODE. Then*

$$\mathbb{E}_{\xi \in \mathcal{R}} [\varphi(t, x_\xi(t))] \sim \ominus \sum_{x \in \mathbb{H}} \varphi(t, x) \rho(t, x).$$

In particular,

$$\mathbb{E}_{\xi \in \mathcal{R}} [\varphi(t, x_\xi(t))] \sim \int \varphi(t, x) \rho(t, x) \Delta x.$$

Proof. We have

$$\begin{aligned} \mathbb{E}_{\xi \in \mathcal{R}} [\varphi(t, x_\xi(t))] &= \frac{1}{|\mathcal{R}|} \sum_{\xi \in \mathcal{R}} \varphi(t, x_\xi(t)) = \frac{1}{|\mathcal{R}|} \sum_{x \in \mathbb{H}} \left[\sum_{x \leq x_\xi(t) < x + \ominus} \varphi(t, x_\xi(t)) \right] \\ &= \frac{1}{|\mathcal{R}|} \sum_{\xi \in \mathcal{R}} \varphi(t, x_\xi(t)) = \frac{1}{|\mathcal{R}|} \sum_{x \in \mathbb{H}} \left[\sum_{x \leq x_\xi(t) < x + \ominus} (\varphi(t, x) + \eta_\xi(t)) \right] \end{aligned}$$

where $\eta_\xi(t) := \varphi(t, x_\xi(t)) - \varphi(t, x)$ is infinitesimal. Hence

$$\frac{1}{|\mathcal{R}|} \sum_{\xi \in \mathcal{R}} \eta_\xi(t) \sim 0$$

and so

$$\begin{aligned} \mathbb{E}_{\xi \in \mathcal{R}} [\varphi(t, x_\xi(t))] &= \frac{1}{|\mathcal{R}|} \sum_{x \in \mathbb{H}} \left[\sum_{x \leq x_\xi(t) < x + \ominus} \varphi(t, x) \right] \\ &= \frac{1}{|\mathcal{R}|} \sum_{x \in \mathbb{H}} (\varphi(t, x) \cdot |\{x_\xi \in \mathcal{S} : x \leq x_\xi(t) < x + \ominus\}|) \\ &= \frac{1}{|\mathcal{R}|} \sum_{x \in \mathbb{H}} [\varphi(t, x) \cdot \ominus |\mathcal{R}| \rho(t, x)] = \ominus \sum_{x \in \mathbb{H}} [\varphi(t, x) \rho(t, x)] \end{aligned}$$

■

Now we can prove Theorem 27

Proof. Chosen an arbitrary $\varphi \in \mathcal{C}([0, 1] \times \mathbb{R})$ bounded in the second variable, we have that

$$\varphi(1, x_\xi(1)) - \varphi(0, x_0) = \ominus \sum_{t \in [0, 1 - \ominus]_{\mathbb{H}}} \frac{\Delta \varphi}{\Delta t}(t, x_\xi(t)),$$

Now we assume that $\varphi \in \mathcal{D}([0, 1] \times \mathbb{R})$. Since x_ξ solves eq. (22), we can apply Theorem 21, and we obtain

$$\begin{aligned} -\varphi(0, x_0) &\sim \ominus \sum_{t \in [0, 1]_{\mathbb{H}}} \left[\varphi_t + \varphi_x \cdot \frac{\Delta x}{\Delta t} + \frac{\ominus}{2} \varphi_{xx} \cdot \left(\frac{\Delta x}{\Delta t} \right)^2 \right] = \\ &\sim \ominus \sum_{t \in [0, 1]_{\mathbb{H}}} \left[\varphi_t + \varphi_x \cdot (f + h\xi) + \frac{\ominus}{2} \varphi_{xx} \cdot (f + h\xi)^2 \right] = \\ &= \ominus \sum_{t \in [0, 1]_{\mathbb{H}}} (\varphi_t + f\varphi_x) + (\varphi_x h + \ominus \varphi_{xx} f) \xi + \frac{\ominus}{2} \varphi_{xx} f + \frac{\ominus}{2} \varphi_{xx} h^2 \xi^2 \end{aligned} \quad (32)$$

Now we want to compute $\mathbb{E}_{\xi \in \mathcal{R}}$ of each piece of the right hand side of the above equation: by lemma 35,

$$\mathbb{E}_{\xi \in \mathcal{R}} [\varphi_t + f\varphi_x] = \ominus \sum_{x \in \mathbb{H}} [\varphi_t + f\varphi_x] \rho \quad (33)$$

for every $t \in [0, 1]$.

Let us consider the second piece: $(\varphi_x h + \ominus \varphi_{xx} f) \xi$. If we set

$$F(t, x_\xi(t)) = \varphi_x(t, x_\xi(t))h(t, x_\xi(t)) + \ominus \varphi_{xx}(t, x_\xi(t))f(t, x_\xi(t))$$

it turns out that $F(t, x_\xi(t))$ is bounded. In fact if $x_\xi(t)$ is bounded, $\varphi_x(t, x_\xi(t))$, $h(t, x_\xi(t))$, $\varphi_{xx}(t, x_\xi(t))$ and $f(t, x_\xi(t))$ are bounded since they are standard functions; if $x_\xi(t)$ is unbounded, $\varphi_x(t, x_\xi(t)) = 0$. Then we can apply lemma 34 and we get that

$$\mathbb{E}_\xi [(\varphi_x h + \ominus \varphi_{xx} f) \xi] \sim 0. \quad (34)$$

Moreover

$$\mathbb{E}_\xi \left[\frac{\ominus}{2} \varphi_{xx} f \right] \sim 0 \quad (35)$$

since $\frac{\ominus}{2} \varphi_{xx} f \sim 0$.

Finally, let us see the last term: $\frac{\ominus}{2} \varphi_{xx} h^2 \xi^2$. We can see that $\frac{\ominus}{2} \varphi_{xx} h^2$ is bounded (actually, it is infinitesimal); then we can apply lemma 34 with $F(t, x_\xi(t)) = \frac{\ominus}{2} \varphi_{xx}(t, x_\xi(t))h(t, x_\xi(t))^2$:

$$\begin{aligned}
\mathbb{E}_\xi \left[\frac{\ominus}{2} \varphi_{xx} h^2 \xi^2 \right] &\sim \alpha \mathbb{E}_\xi \left[\frac{\ominus}{2} \varphi_{xx} h^2 \right] \\
&= \frac{1}{2} \mathbb{E}_\xi [\varphi_{xx} h^2] \\
&= \frac{1}{2 |\mathcal{R}|} \sum_{\xi \in \mathcal{R}} \varphi_{xx}(t, x(t)) h(t, x(t))^2
\end{aligned} \tag{36}$$

Then, by lemma 35

$$\mathbb{E}_\xi \left[\frac{\ominus}{2} \varphi_{xx} h^2 \xi^2 \right] \sim \frac{\ominus}{2} \sum_{x \in \mathbb{H}} \varphi_{xx}(t, x) h(t, x)^2 \rho(t, x)$$

Then, by (32),..., (36)

$$\begin{aligned}
-\varphi(0, x_0) &= \mathbb{E}_{t, \xi} [-\varphi(0, x_0)] \\
&\sim \ominus \sum_{t \in [0, 1]_{\mathbb{H}}} \left(\mathbb{E}_\xi [\varphi_t + f \varphi_x] + \mathbb{E}_\xi [(\varphi_x h + \ominus \varphi_{xx} f) \xi] \right) + \\
&\quad + \ominus \sum_{t \in [0, 1]_{\mathbb{H}}} \left(\mathbb{E}_\xi \left[\frac{\ominus}{2} \varphi_{xx} f \right] + \mathbb{E}_\xi \left[\frac{\ominus}{2} \varphi_{xx} h^2 \xi^2 \right] \right) \\
&\sim \ominus^2 \sum_{t \in [0, 1]_{\mathbb{H}}} \left(\sum_{x \in \mathbb{H}} (\varphi_t + f \varphi_x) \rho + \varphi_{xx} h^2 \rho \right) \\
&\sim \int \int (\varphi_t + f \varphi_x + \varphi_{xx} h^2) \rho \, dx dt
\end{aligned}$$

Since the first and the last term of this equation are standard, we get eq. (26). ■

5 Conclusion

In this section we will make some comments suggested by the results of this paper.

The first comment is that,

- *in many applications of Nonstandard Analysis, only elementary facts and techniques of nonstandard calculus seem to be necessary.*

In fact α -theory is much simpler than the usual Nonstandard Analysis, but it seems absolutely adequate to treat the kind of problems considered here.

However, if you compare this paper with [2], [8] and [11], the reason why this paper is much simpler lies in the fact that we have not made a nonstandard theory of the *stochastic differential equation*, but rather, we have replaced them with the *stochastic grid equations* which are much simpler mathematical objects; in this contest the Ito integral is replaced by the α -integral and the proof of the key point of this theory, the Ito formula, reduces to an exercise.

The reason why a much simpler object can represent the processes of diffusion at microscopic level is that we have taken the infinitesimals seriously and we have used them to model an aspect of the "physical reality". We think that in general,

- *the advantages of a theory which includes infinitesimals rely more on the possibility of making new models rather than in the demonstration techniques.*

In the case considered in this paper, the way the model has been constructed appears quite natural: the *stochastic grid equations*, which need the notion of infinitesimals, describe the diffusion processes at microscopic level; the Fokker-Plank equation describes the diffusion processes at the macroscopic level and uses standard differential equations (and the theory of distributions when the data are not regular).

The connection of these two levels is given by eq. (8) which relates grid function (microscopic level) with distributions (macroscopic level).

A final remark concerns the *theory of probability*. We have not used the notion of probability to show that every thing can be kept to a very elementary level and that our variant of the Ito formula makes sense also in a context where probability does not appear.

However probability can be introduced in a very elementary way. We may think of the stochastic class \mathcal{R} as a sample space. The events are the hyperfinite sets $E \subset \mathcal{P}(\mathcal{R})^*$ and the probability P of an event is given by

$$P(E) = \frac{|E|}{|\mathcal{R}|}. \quad (37)$$

For example E might be the event that at a time t_0 , the moving particle lies in the interval $[a, b]$, namely

$$E = \{x_\xi \in \mathcal{S} : x_\xi(t_0) \in [a, b]_{\mathbb{H}}\}$$

In this case we have that

$$P(E) \sim \int_a^b \rho(t, x) \Delta x$$

Obviously, this is the most natural extension to infinite sample spaces of the classical definition of probability. There is no doubt that definition (37), is the most simple and intuitive definition of probability. The price which you pay is that P takes its values in \mathbb{Q}^* and not in \mathbb{R} ; thus, a probability space

(\mathcal{R}, P) is an internal object and, if it is infinite, it is not standard. The problem arise if you want to connect the nonstandard world with the standard one. This operation can be done in a very elegant way via the Loeb integral (see [9] or also [8]). A different way to make easy the connection between this two worlds has been proposed by Nelson [11]. However, we think that, in most cases, it is not necessary to make this connection. Only at the very end you may consider the shadow of the numbers which you have obtained. At least this can be done in the theory which we have exposed in this paper. So, if we accept a mathematical description in which the infinitesimal exist,

- *the probability of an event is given by a hyperrational number,*

and many theorems become simpler. This scheme avoids also some facts in probability theory which are in contrast with the common sense: for example the fact that the union of impossible events gives a possible one (the probability of a non-denumerable set might be non-null even if the probability of each singleton is null).

Concluding the final remark of this paper could be the following one:

- *the infinitesimals should be taken seriously.*

References

- [1] S. ALBEVERIO, J. E. FENSTAD, R. HOEGH-KROHN, AND T. LINDSTRØM, *Non-standard Methods in Stochastic Analysis and Mathematical Physics*, Academic Press, New York, 1986.
- [2] ANDERSON, ROBERT M. A nonstandard representation for Brownian motion and Itô integration. *Bull. Amer. Math. Soc.* 82 (1976), no. 1, 99–101.
- [3] BENCI, V. A Construction of a Nonstandard Universe, in *Advances in Dynamical System and Quantum Physics*, S. Albeverio, R. Figari, E. Orlandi, A. Teta ed., (Capri, 1993), 11–21, World Scientific, (1995).
- [4] BENCI, V., An algebraic approach to nonstandard analysis, in *Calculus of Variations and Partial Differential Equations*, G. Buttazzo, A. Marino and M.K.V. Murthy, eds., Springer-Verlag, 1999, 285–326.
- [5] BENCI, V., DI NASSO, M. Alpha-theory: an elementary axiomatics for nonstandard analysis. *Expo. Math.* 21 (2003), no. 4, 355–386.
- [6] BENCI, VIERI; DI NASSO, MAURO A ring homomorphism is enough to get nonstandard analysis. *Bull. Belg. Math. Soc. Simon Stevin* 10 (2003), no. 4, 481–490.

- [7] BENCI, VIERI; FORTI, MARCO; DI NASSO, MAURO The eightfold path to nonstandard analysis. Nonstandard methods and applications in mathematics, 3–44, Lect. Notes Log., 25, Assoc. Symbol. Logic, La Jolla, CA, 2006.
- [8] KEISLER, H. J. An infinitesimal approach to stochastic analysis. Mem. Amer. Math. Soc. 48 (1984), no. 297, x+184 pp.
- [9] LOEB, PETER A. Conversion from nonstandard to standard measure spaces and applications in probability theory. Trans. Amer. Math. Soc. 211 (1975), 113–122.
- [10] RAGO, E. Una deduzione dell’equazione di Fokker-Planck con metodi Nonstandard, Thesis, University of Pisa, (2001).
- [11] NELSON, E. *Radically elementary probability theory*, Princeton University Press, Princeton, New Jersey, (1987).